UNIT-ROOT, COINTEGRATION AND GRANGER CAUSALITY TEST RESULTS FOR EXPORT AND GROWTH IN OECD COUNTRIES

KÓNYA, László*

Abstract

This article is a supplement to Kónya (2004) which investigates the possibility of the export-led growth and growth-driven export hypotheses by testing for Granger causality between the logarithms of real exports and real GDP in twenty-five OECD countries. In Kónya (2004) two complementary testing strategies were applied. First, depending on the time series properties of the data, causality was tested with Wald tests within finite-order vector autoregressive (VAR) models in levels and/or in first-differences. Then, with no need for pre-testing, a modified Wald (MWald) procedure was used in augmented level VAR systems. For brevity, the results of the unit root, cointegration and MWald tests were not reported in Kónya (2004). The aim of the current article is to fill in this gap.

JEL classification: C12, C22, C32, F43, O57
Keywords: unit root, cointegration, causality, economic growth, export, OECD countries

1. Introduction

This article is a supplement to Kónya (2004) which investigates the possibility of the export-led growth and growth-driven export hypotheses by testing for Granger causality between the logarithms of real exports and real GDP in twenty-five OECD countries. In Kónya (2004) two complementary testing strategies were applied. First, depending on the time series properties of the data causality was tested with standard Wald tests

* Dr László Kónya, Department of Economics and Finance, La Trobe University, Bundoora, VIC 3086, Australia. E-mail: L.Konya@latrobe.edu.au

within finite-order vector autoregressive (VAR) models in levels and/or in first-differences. The disadvantage of this strategy is that the final outcome might heavily depend on preliminary test results which, themselves, are often uncertain and misleading. Then, in order to reduce the impact of pre-testing on the conclusions regarding causality, the modified Wald (MWald) test of Toda and Yamamoto (1995) for Granger causality was also used in augmented level VAR systems. The advantage of this procedure is that it is valid even under uncertainty about integration and cointegration.[1]

For brevity, not all test results were reported in Kónya (2004). The aim of the current paper is to fill in this gap by reporting, comparing and interpreting the outcomes of various unit-root, cointegration and MWald tests. As a by-product, this paper also illustrates that different methods and alternative specifications can and often do lead to contradicting conclusions. The rest of the article is organized as follows. The data and its univariate properties, with special regard to the unit-root and stationarity tests, are discussed in Section 2. The no-cointegration and cointagration test results are summarised in Section 3. Section 4 discusses the MWald test results. The concluding remarks can be read in Section 5. Finally, the time-series plots of the data, for six countries, can be seen in the Appendix.

2. The data and its univariate properties

All data utilised in this study are from *EconData, World Bank World Tables*. The data set comprises annual measures on 25 OECD countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Greece, Hungary, Iceland, Ireland, Italy, Japan, Korea Rep., Luxembourg, Mexico, Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, Switzerland, UK and USA. The sample period is 1960-1997 for all countries, except Hungary (1970-1998), Korea and Mexico (1960-1998). The variables are GDP in 1995 $US million (GDP), exports of goods and services in 1995 $US million (EXP), imports of goods and services in 1995

[1] Dolado and Lütkepohl (1996) propose a similar technique.

$US million (IMP) and openness (OPEN), defined as (EXP+IMP)/GDP.[2] GDP, EXP and OPEN have been transformed in natural logarithms and the resulting variables are denoted as LNGDP, LNEXP and LNOPEN.

Prior to testing for Granger causality it is important to establish the properties of the time series involved. In particular, the order of integration and the existence of common trends are of major importance. The LNGDP and LNEXP data series are shown in the Appendix (*Figures* 1-24). A simple visual inspection of the figures suggests that the data series are fairly smooth but trended. However, it is impossible to tell without further analysis whether the trend components are deterministic or stochastic, and whether the LNGDP and LNEXP series have any common trend. We perform our analysis in two steps. First, in this section, we test for unit roots vs. stationarity. Then, in the following section, we test for no-cointegration vs. cointegration. In each test the significance level is 10 percent.

Testing for a (non-seasonal) unit root is usually based on the following first-order autoregressive, AR(1), model:

$$y_t = a_1 y_{t-1} + e_t$$

where e_t is supposed to be a white-noise error and the initial value, y_0, is constant. This basic model, with no deterministic terms, will be referred to as *Model 1*. It can be augmented with a constant term, called drift, to obtain *Model 2*

$$y_t = a_0 + a_1 y_{t-1} + e_t$$

[2] The original trade figures extracted from the World Bank=s *World Tables* are nominal exports and imports given in $US, and index numbers (1995=100) calculated from real exports and imports in 1995 local currencies. In order to obtain real trade data in 1995 $US, the index series have been multiplied by the 1995 trade figures given in 1995 $US (million).

and with a deterministic trend to obtain *Model 3*

$$y_t = a_0 + a_1 y_{t-1} + a_2 t + e_t$$

In each of these models the most crucial parameter is a_1. When $|a_1| < 1$ and t is sufficiently large, the expected value of y_t is zero in *Model 1*, constant in *Model 2*, and a linear function of t in *Model 3*. However, if $a_1 = 1$, that is the underlying AR(1) process has a unit root, then the expected value of y_t is constant in *Model 1*, a linear function of t in *Model 2*, and a quadratic function of t in *Model 3.* As regards the second moments, the variance and autocovariances of y_t are time invariant, at least for sufficiently large values of t, if $|a_1| < 1$, but they depend on t, even asymptotically, if $a_1 = 1$. In particular, for $a_1 = 1$, the variance of y_t becomes a linear function of t, while the autocovariances are linear functions of the time lag. Hence, the AR(1) processes described by models (1)-(3) are, at least asymptotically, (weakly) stationary around some constant or linear time trend if $|a_1| < 1$, but are integrated of order one, I(1), and thus non-stationary, if $a_1 = 1$.

Assuming that the true data generating processes (DGP) are indeed AR(1) processes, the trended LNGDP and LNEXP series might have been generated either by stationary processes with a deterministic trend component, or by unit-root processes with a drift. For this reason, we test for unit roots first allowing for linear trends and then for only drifts, but occasionally we experiment with quadratic trends, as well. The literature on modern time-series analysis offers dozens of methods for unit-root testing. Since each of them has some weakness, in empirical studies it is better not to rely on any particular test but to use several of them. To this end, we start with three unit root/stationarity tests advocated by Dickey and Fuller (1979, 1981), Park and Fuller (1995) and Kwiatkowski *et al.* (1992). From hereon we refer to these tests as DF, PF and KPSS, respectively.[3] In spite of its relatively low power, the (augmented) DF t-test is undoubtedly the most popular unit-root test, so it is used as a benchmark procedure. The (augmented) weighted symmetric PF t-test is a

[3] A useful discussion of unit-root testing can be found in Maddala and Kim (1998, Ch. 3, 4).

modification of the DF t-test and, in general, it is more powerful and performs better. The KPSS test, unlike the DF and PF tests, takes stationarity as the null hypothesis.4 It is a special case of a more general Lagrange Multiplier (LM) test for parameter constancy against the alternative hypothesis that a parameter follows a random walk. Unfortunately, the KPSS test also suffers from low power, yet it can prove to be useful for confirmatory analysis in conjunction with the other two unit root tests. If both types of tests lead to the same conclusion, one may have more confidence in the outcomes.[5] The results of the DF, PF and KPSS tests on LNGDP and LNEXP are summarised in columns 2-4 of Tables 1 and 2.

As a second step, we also consider the possibility of double unit roots. The reason for doing so is that for many countries the LNGDP and LNEXP series are quite smooth, and just as integrated series of order one, I(1), tend to be smoother than stationary series, I(0), integrated series of order two, I(2), tend to be even smoother than I(1) series. This is not simply a statistical issue, I(2) variables have an important economic

---

[4] This is an important point since the non-rejection of a null hypothesis should never grant its acceptance. In relation to the DF and PF tests this means that a non-rejected unit-root null hypothesis is a relatively weak outcome and it must not be considered a strong support for the presence of a unit root.

[5] The DF and PF tests have been performed by the COINT command of TSP 4.5. This command selects the optimal number of augmenting lags (*mlag*) on the basis of the AIC2 rule. According to this rule, *mlag* = *j*+2, where *j* is the number of lags which minimizes the Akaike Information Criterion. Since this lag selection method does not necessarily assure white-noise residuals, we re-estimated the augmented DF regressions with OLS and tested the residuals for autocorrelation of orders 1-4 using the Breusch-Godfrey LM and the Ljung-Box portmanteau tests. When it was necessary, *mlag* selected by TSP has been increased (up to a maximum of 8 lags). The KPSS test requires the estimation of the long-run variance which depends on the lag truncation parameter, l. If there is no need to correct for autocorrelation at all, then l is zero. Otherwise several positive l values have to be considered in order to check how robust the test outcomes are for the choice of lag truncation. For this reason, we used l = 3, 5 and 7. As for the deterministic term(s) included in the AR model, we experimented with drift, linear and quadratic trends and chose the simplest specification leading to the maintenance or rejection of H_0.

interpretation, as well. Namely, if a log-transformed series is I(2), then its first difference, i.e. the growth rate, is a random walk, and the second difference, which is the change in the growth rate, is stationary.[6]

We apply two tests for double unit roots, recommended by Dickey and Pantula (1987) and Hasza and Fuller (1979). The first one, denoted as DP, is a sequential test procedure which starts with the supposedly highest order of integration and tests it down gradually. Assuming, for example, that there are at most two unit roots, first I(2) is tested against I(1) and then, provided that the null hypothesis is rejected, I(1) is tested against I(0). The second test, HF, is a single-step method which compares the null hypothesis of I(2) to the composite alternative hypothesis that the series is I(0), I(1) or explosive. The DP test is more powerful against the I(1) alternative than HF and, due to the more informative alternative hypotheses, it is usually more useful. However, it can be misleading if the DGP is explosive since in small samples explosive processes can behave quite similarly to I(2) processes. For this reason, we apply DP with various sets of deterministic terms on all series and when I(2) is maintained, we use HF, as well.[7] If HF rejects the null, then following Haldrup (1998, p. 608), we reconsider the DF *t*-test statistic to check whether it is significantly positive, indicating an explosive alternative.[8] The results of the DP and HF tests on LNGDP and LNEXP are summarised in columns 5 and 6 of Tables 1 and 2.

As it turns out, different unit root/stationarity tests on the same specification, and a given test on different specifications often lead to conflicting outcomes. In fact, there is more or less unambiguity in only one third of the countries (Australia, Belgium, Iceland, Ireland, Luxembourg, Norway, Spain and the UK).[9] In all other cases we follow

---

[6] About the analysis of I(2) variables see e.g. Maddala and Kim (1998, Ch. 11) and Haldrup (1998).

[7] DP could safely reject I(3) in favour of I(2) for all series.

[8] In the DP and HF tests *mlag* (=1,..., 8) was selected as the shortest lag structure ensuring white noise residuals. When autocorrelation could not be eliminated we chose the lag structure generating the >whitest= residuals.

[9] This finding is not underlines

two simple rules: a more parsimonious specification is preferred to a less parsimonious one, and a rejected null hypothesis is favoured against a maintained one. When these rules are not satisfactory, we do not strive to arrive at a simple conclusion, but leave both options alive.

Taking all results together, it is not plausible for any country to exclude the possibility of a single unit root, neither for LNGDP nor for LNEXP. If these variables are indeed I(1), then they might or might not have a common stochastic trend, i.e. they might or might not be cointegrated. This is the next issue to be resolved.

3. Testing for no cointegration/cointegration

Similarly to unit-root testing, the econometric literature offers many different cointegration tests. We apply three of them, recommended by Engle and Granger (1987) [EG], Johansen and Juselius (1990) [JJ] and Shin (1994) [S].[10] The EG test is a multivariate generalisation of the DF *t*-test with the null hypothesis of no-cointegration, while the S test addresses the cointegration hypothesis directly, in the vein of the KPSS methodology. Both EG and S are two-step, single-equation, residual-based tests.

Consequently, in spite of the fact that in a cointegrating relationship the variables have equal status, the actual outcomes of these tests might be affected by the choice of the dependent variable in the cointegrating regression. On the other hand, the JJ likelihood ratio trace test is a system method based on vector autoregression (VAR). However, in the case of low-order VAR models or small samples ($n < 100$) this test is seriously biased toward spuriously detecting cointegration. Moreover, for practical purposes, the JJ test is not a single step method either, and incorrect model specification by the researcher can lead to misleading results.

[10] In addition to the original sources, the interested reader can find a resourceful summary of these and several other cointegration tests in Maddala and Kim (1998, Ch. 5 and 6). About some of the pitfalls of these tests see also Guisan (2001).

Table 1. Summary of the unit root/stationarity test results on LNGDP(Number of unit roots, required deterministic terms)

Country	DF	WS	KPSS	DP	HF	Conclusion
Australia	1	1	1	1		I(1)
Austria	0 or 1TR	1	1	2	0 or 1, DR	I(0) DR or I(1) TR
Belgium	1	1	1	1TR		I(1) TR
Canada	0 or 1TR	1	1	0 or 1TR		I(0) or I(1) TR
Denmark	0TR or 1	1	1	1TR		I(1) TR
Finland	0 or 1TR	1	1	0QTR or 1		I(0) QTR or I(1)
France	0 or 1TR	1	1	1TR		I(0) or I(1) TR
Greece	0 or 1TR	1	1	1TR		I(0) or I(1) TR
Hungary	1	1	1	2	0 or 1, DR	I(1) DR
Iceland	1	1	1	1		I(1)
Ireland	1	1	0TR or 1	1		I(1)
Italy	0 or 1TR	1	1	0 or 1TR		I(0) or I(1) TR
Japan	0 or 1TR	1	1	2	0 or 1, DR	I(0) DR or I(1) TR

Korea Rep.	1	1	0TR or 1	1		I(1)
Luxembourg	1	1	1	1		I(1)
Mexico	0 or 1TR	1	1	0 or 1TR		I(0) or I(1) TR
Netherlands	1	1	1	2	0 or 1, DR	I(1) DR
New Zealand	1	1	1	1		I(1)
Norway	1	1	1	1		I(1)
Portugal	0 or 1TR	1	1	0 or 1TR		I(0) or I(1) TR
Spain	1	1	1	1		I(1)
Sweden	0	1	1	0 or 1TR		I(0) or I(1) TR
Switzerland	1	1	1	1		I(1)
UK	1	1	0TR or 1	1		I(1)
USA	0TR or 1	1	1	1		I(0) TR or I(1)

Note: a) DR, TR and QTR refer to drift, linear trend and quadratic trend components, respectively. b) The items in bold are test outcomes generated by rejected null hypotheses.

Table 2. Summary of the unit root/stationarity test results on LNEX(Number of unit roots, required deterministic terms)

Country	DF	WS	KPSS	DP	HF	Conclusion
Australia	1	1	0TR or 1	1		I(1)
Austria	1	1	1	1		I(1)
Belgium	1	1	1	1		I(1)
Canada	1	1	1	1		I(1)
Denmark	0TR or 1	1	1	0TR or 1		I(0) TR or I(1)
Finland	1	1	1	1		I(1)
France	1	1	1	1		I(1)
Greece	1	1	1	1		I(1)
Hungary	1	1	1	2	0 or 1, DR	I(1) DR
Iceland	1	1	1	1		I(1)
Ireland	1	1	1	1		I(1)
Italy	1	1	1	1		I(1)
Japan	1	1	1	1TR		I(1) TR
Korea Rep.	0 or 1TR	1	1	0		I(0) or I(1) TR
Luxembourg	1	1	0TR or 1	1		I(1)
Mexico	1	1	0TR or 1	1		I(1)

Netherlands	1	1	1	1		I(1)
New Zealand	0TR or 1	0TR or 1	0TR or 1	1		I(0) TR or I(1)
Norway	1	1	0TR or 1	1		I(1)
Portugal	1	1	0TR or 1	0QTR		I(0) QTR or I(1)
Spain	1	1	0TR or 1	1		I(1)
Sweden	1	1	0TR or 1	1		I(1)
Switzerland	0 or 1TR	1	1	1TR		I(0) or I(1) TR
UK	1	1	1	2	0 or 1, DR	I(1) DR
USA	1	1	0TR or 1	1		I(1)

Note: a) DR, TR and QTR refer to drift, linear trend and quadratic trend components, respectively. b) The items in bold indicate test outcomes generated by rejected null hypotheses.

In order to avoid spurious results, we consider alternative specifications, with and without deterministic terms. We start with time polynomials of the highest order occurring in the likely DGPs behind LNGDP and LNEXP. Lower orders, however, might be also sufficient, granted that the deterministic components present in the two series eliminate each other. On each specification EG and S were performed twice, using both

LNGDP and LNEXP in turn as dependent variable.[11] The results of the EG, JJ and S tests are summarised in *Table 3.*

Like the unit-root test results, in most cases the cointegration test results are also ambiguous. There are only two countries (Iceland and the Netherlands) for which all three tests support cointegration, CI(1,1), and four countries (Australia, Belgium, Canada and Hungary) for which they all indicate no-cointegration. For the other nineteen countries the results are contradictory. In six cases (Finland, Ireland, New Zealand, Norway, Switzerland and the USA), the EG test does not reject no-cointegration, but the S test rejects cointegration and the JJ test rejects no-cointegration. In two cases (Japan and Spain), either the EG or the JJ test rejects no-cointegration while the other does not reject it, and the S test does not reject cointegration. In five cases (Austria, Denmark, France, Luxembourg, Portugal and Sweden) the JJ test detects two cointegrating vectors suggesting that, contrary to what we concluded from the unit-root/stationarity test results, LNGDP and LNEXP are in fact stationary. Finally, in five cases (Greece, Italy, Korea, Mexico and the UK), neither test can reject the corresponding null hypothesis. The fact that the EG, JJ and S tests lead to the same conclusion in only six countries out of 25 reconfirms how unreasonably it is to test for unit roots or for cointegration with only one technique and, upon the results, enforcing one particular DGP on the observed time series. Allowing for alternative DGPs might often be a more reasonable approach.

In summary, the revealed time series properties of the data imply the following four cases:

a. LNGDP and LNEXP are both I(0). There are seven countries where this possibility cannot be safely excluded: Austria, Denmark, France, Luxembourg, Portugal and Sweden.
b. One of the series is I(0) and the other one is I(1). This is a plausible option for Austria, Canada, Denmark, Finland, France,

[11] In the EG test the lag structure was selected in the same way than in the DP and HF tests, in the JJ test the lag structure was based on AIC, while in the S test, we have experimented with three different lag lengths: 2, 3 and 4.

Greece, Italy, Japan, Korea, Mexico, New Zealand, Portugal, Sweden, Switzerland and the USA.

c. Both series are I(1) but they are not CI(1,1). This might be true for Australia, Belgium, Canada, Finland, Greece, Hungary, Ireland, Italy, Korea, Luxembourg, Mexico, New Zealand, Norway, Switzerland, UK and the USA.

d. LNGDP and LNEXP are CI(1,1). This possibility cannot be overlooked in the case of Denmark, Finland, France, Greece, Iceland, Ireland, Italy, Japan, Korea, Mexico, the Netherlands, New Zealand, Portugal, Spain, Sweden, Switzerland, UK and the USA.

4. Testing for Granger causality with the MWald test

The concept of Granger causality, by which we actually understand *precedence*, is based on the idea that a cause cannot come after its effect.[12] More precisely, variable X is said to Granger-cause another variable, Y, if the current value of Y (y_t) is conditional on the past values of X (x_{t-1}, x_{t-2}, ... , x_0) and thus the history of X is likely to help predict Y.
Note, that this is causality for one period ahead. This concept is generalized by Dufour and Renault (1998) to causality h periods ahead, and to causality up to horizon h, where h is a positive integer that can be infinite. They show that in a bivariate system no-causality for one period ahead implies no-causality at, or up to, any horizon.

This is a clear advantage of a bivariate system over a trivariate system, (X, Y, Z) for example, where causality between X and Y can arise via the auxiliary variable Z. Namely, X might cause Z one period ahead, which in turn might cause Y at a subsequent period. This indirect, two-period ahead causality might exist even if there is no direct, one-period ahead causality between X and Y. However, if there is no causality between X and Y for two periods ahead then there is no causality between them at, or up to, longer horizons either. This difference between bivariate and trivariate systems implies that they require different strategies to test for causality at

[12] From now on, >causality= always refers to >Granger causality=.

horizons beyond one period.

Table 3. Summary of the no-cointegration/cointegration test results on LNGDP and LNEX

Country	Model	AEG	S	JJ	Conclusion
Australia	1, 2	not CI(1,1)	not CI(1,1)[3]	not CI(1,1)	not CI(1,1)
Austria	3	not CI(1,1)	CI(1,1)[2]	inconclusive	contradicting
Belgium	3	not CI(1,1)	not CI(1,1)	not CI(1,1)	not CI(1,1)
Canada	3	not CI(1,1)	not CI(1,1)	not CI(1,1)	not CI(1,1)
Denmark	3	CI(1,1)	CI(1,1)	inconclusive	CI(1,1)
Finland	1, 2	not CI(1,1)[1]	not CI(1,1)[3]	CI(1,1)	contradicting
France	3	not CI(1,1)	CI(1,1)	inconclusive	contradicting
Greece	3	not CI(1,1)	CI(1,1)	not CI(1,1)	contradicting
Hungary	2	not CI(1,1)	not CI(1,1)	not CI(1,1)	not CI(1,1)
Iceland	1, 2	CI(1,1)[4]	CI(1,1)	CI(1,1)	CI(1,1)
Ireland	1, 2	not CI(1,1)	not CI(1,1)[3]	CI(1,1)[3]	contradicting
Italy	3	not CI(1,1)	CI(1,1)	not CI(1,1)	contradicting
Japan	2, 3	CI(1,1)[4]	CI(1,1)	not CI(1,1)	CI(1,1)

Korea Rep.	3	not CI(1,1)[1]	CI(1,1)[2]	not CI(1,1)	contradicting
Luxembourg	1, 2	not CI(1,1)	not CI(1,1)	not CI(1,1)	not CI(1,1)
Mexico	3	not CI(1,1)	CI(1,1)[2]	not CI(1,1)	contradicting
Netherlands	2	CI(1,1)	CI(1,1)	CI(1,1)	CI(1,1)
New Zealand	1, 2	not CI(1,1)[1]	not CI(1,1)[3]	CI(1,1)[3]	contradicting
Norway	1, 2	not CI(1,1)	not CI(1,1)[3]	CI(1,1)[3]	contradicting
Portugal	3	not CI(1,1)	CI(1,1)[2]	inconclusive	contradicting
Spain	1, 2	not CI(1,1)[1]	CI(1,1)	CI(1,1)[3]	contradicting
Sweden	3	not CI(1,1)	CI(1,1)	inconclusive	contradicting
Switzerland	3	not CI(1,1)	not CI(1,1)	CI(1,1)	contradicting
UK	2	not CI(1,1)	CI(1,1)[2]	not CI(1,1)	contradicting
USA	1, 2	not CI(1,1)	not CI(1,1)[3]	CI(1,1)[3]	contradicting

Note: a) *Model 1* has no deterministic term, *Model 2* has an intercept term, while *Model 3* has a linear trend. b) [1]: There is some sign of cointegration. [2]: There is some sign of no-cointegration. [3]: Only for Model 1. [4]: Only for Model 2. c) The items in bold indicate test outcomes generated by rejected null hypotheses.

In this section we study the possibility of Granger causality between

LNGDP and LNEXP. We start with a simple bivariate system, but later we augment the information set with a third variable, LNOPEN. This third variable is treated as auxiliary in the sense that it is not directly involved in the causality test. Therefore, the possibility of indirect causality is ignored, though two-step causality may arise when LNGDP cause LNEXP and/or LNEXP cause LNGDP indirectly, via LNOPEN. Our focus remains on direct causality between LNGDP and LNEXP. It is important to remember this limitation of the analysis when comparing the bivariate and trivariate results to each other.[13]

In order to re-enforce the Granger-causality test results, we apply two complementary strategies. The first one, let us call it indirect approach, assumes that the variables are stationary or can be made stationary by differencing. It makes use of pre-testing for unit roots and cointegration and, depending on the outcomes, testing for causality is undertaken within VAR models of different specifications. When both series are deemed I(0), a VAR model in levels is used. When one of the series is found I(0) and the other one I(1), VAR is specified in the level of the I(0) variable and in the first difference of the I(1) variable. When both series are determined I(1) but not cointegrated, the proper model is VAR in terms of the first differences. Finally, when the series are cointegrated, we can use a vector error correction (VECM) model or, for a bivariate system, a VAR model in levels. Obviously, the weakness of this strategy is that incorrect conclusions drawn from preliminary analyses might be carried over onto the causality tests. In the light of the unit-root and cointegration test results, this possibility must be taken seriously. The ambiguities of pre-testing might have great impact on the final conclusions regarding Granger-causality, unless different VAR specifications lead to the same results. The second strategy, let us call it direct approach, is free of this problem. It is based on the procedure of Toda and Yamamoto (1995) which does not rely so heavily on pre-testing, though some knowledge of the maximum order of integration and of the lag structure is still required.

[13] The typically applied methods for Granger causality do not consider indirect effects for two or more periods ahead. Rare exceptions are the two strategies suggested by Giles (2000) which can be applied to test for two-step causality in trivariate VAR systems with an auxiliary variable.

The indirect approach is based on the following VAR system:

$$
\begin{aligned}
y_t &= a_1 + \sum_{i=1}^{mlag} b_{1i} y_{t-i} + \sum_{i=1}^{mlag} g_{1i} x_{t-i} + e_{1t} \\
x_t &= a_2 + \sum_{i=1}^{mlag} b_{2i} y_{t-i} + \sum_{i=1}^{mlag} g_{2i} x_{t-i} + e_{2t}
\end{aligned}
\qquad (1)
$$

where it is assumed that y_t and x_t are stationary, i.e. they denote the levels or the (first) differences of LNGDP and LNEXP, respectively; and that e_{1t} and e_{2t} are white-noise disturbances. If necessary, (1) can be augmented with a deterministic linear or quadratic trend, and additional auxiliary variables might be also considered.

On the basis of the preliminary test results, we estimated (1) with various deterministic trends and

i. with X = LNEXP, Y = LNGDP for Australia, Austria, Denmark, France, Iceland, Ireland, Italy, Japan, Korea, Luxembourg, Mexico, the Netherlands, Norway, Portugal, Spain, Sweden and the UK.;

ii. with X = ?LNEXP, Y = LNGDP for Denmark, Korea, New Zealand, Portugal and Switzerland, and with X = LNEXP, Y = ?LNGDP for Austria, Canada, Finland, France, Greece, Italy, Japan, Mexico, Portugal, Sweden and the USA.[14]

iii. with X = ?LNEXP, Y = ?LNGDP for Australia, Belgium, Canada, Finland, Greece, Hungary, Ireland, Italy, Korea, Luxembourg, Mexico, New Zealand, Norway, Switzerland, the UK and the USA.

In each case we experimented with four different lag lengths (*mlag* = 1, 2, 3, 4) over the same sample. The $?_{1,1} = \ldots = ?_{1,mlag} = 0$ and $ß_{2,1} = \ldots = ß_{2,mlag} = 0$ parameter restrictions were tested with standard Wald tests which,

[14] ? denotes the first-difference operator.

under the null hypotheses, have a limiting ?^{2} distribution with *mlag* degrees of freedom. Since the results have already been reported in Kónya (2004), they are not repeated here. We shall rather focus on the direct approach.

The test proposed by Toda and Yamamoto (1995) is a modified Wald (MWald) test for linear restrictions on some parameters of an augmented VAR(*mlag* + *d*) in levels, where *d* is the highest order of integration suspected in the system, usually at most two. In the bivariate case this model without deterministic terms can be written as

$$
\begin{aligned}
y_t &= a_1 + \sum_{i=1}^{mlag} b_{1i} y_{t-i} + \sum_{i=mlag+1}^{mlag+d} b_{1i} y_{t-i} + \sum_{i=1}^{mlag} g_{1i} x_{t-i} + \sum_{i=mlag+1}^{mlag+d} g_{1i} x_{t-i} + e_{1t} \\
x_t &= a_2 + \sum_{i=1}^{mlag} b_{2i} y_{t-i} + \sum_{i=mlag+1}^{mlag+d} b_{2i} y_{t-i} + \sum_{i=1}^{mlag} g_{2i} x_{t-i} + \sum_{i=mlag+1}^{mlag+d} g_{2i} x_{t-i} + e_{2t}
\end{aligned}
\tag{2}
$$

In this case the test statistic does not depend on any nuisance parameter and under the null hypothesis it has an asymptotic ?^{2} distribution with the usual degrees of freedom, granted that *d*# *mlag*. The last *d* lags are not considered explicitly in the Wald test. These extra lags, however, are necessary to ensure the asymptotically ?^{2} sampling distribution of the test statistic.

This test procedure has three advantages. Firstly, it can be used in possible integrated and cointegrated systems, without pre-testing for cointegration. Secondly, Rambaldi and Doran (1996) have shown that computationally the MWald test is very simple, since it can be run in a seemingly unrelated regression. Thirdly, according to the Monte Carlo experiments on bivariate and trivariate models performed by Zapata and Rambaldi (1997), despite the intentional over-fitting, the MWald test performs as well as similar but more complicated test procedures in samples of size fifty at least. This time, unfortunately, we work with shorter time series, and the extra, redundant regressors may lead to costly losses in power and efficiency. Nevertheless, it is worth to apply this procedure and compare the outcomes to the results obtained via the indirect approach.

First we used the MWald test on the LNGDP-LNEXP bivariate system. As for deterministic terms, (2) was occasionally augmented with a linear or quadratic trend. We assumed that the maximal order of integration is one, i.e. $d = 1$, and experimented with $mlag + d = 2, 3, 4, 5$. For each country the preferred *mlag* value was selected on the basis of AIC, SBIC statistics from VAR(*mlag*) estimated by OLS over the same sample. When these statistics chose different *mlag* values, preference was given to the one which produced non-autocorrelated, or at least 'whiter', residuals. If, however, the conclusions proved to be sensitive to the lag structure, we report the more stable outcome, i.e. the one supported by more lag lengths.[18]

The results, shown in Table 4, suggest one-way causality from LNEXP to LNGDP in the case of Australia (*Model 2*), Austria (*Model 2*), Belgium, Hungary (*Model 3*), Iceland, Ireland, Spain and Switzerland, from LNGDP to LNEXP in Canada, Finland (Model 4), Italy, Japan, Korea, Mexico (Model 2), New Zealand (Model 3), Portugal and the USA (Model 2), and two-way causality seems to be likely in Denmark (Model 3), Finland (Models 2, 3), New Zealand (Model 2), Sweden and the UK.

The Model 2 - VAR($mlag+d$) with constant;
The Model 3 - VAR($mlag+d$) with linear trend;
The Model 4: VAR($mlag+d$) with quadratic trend.
And, the Causality Test is: H_{01} : LNEXP does not cause LNGDP;
H_{02} : LNGDP does not cause LNEXP.

[18] Unfortunately, for some countries the causality test results are so sensitive to the lag structure that, despite all care, the conclusions drawn from the MWALD tests are certainly subjective and thus disputable. This is especially true for New Zealand and Portugal.

TABLE 4: Mwald tests for Granger causality - bivariate VAR (mlag + d) in levels with d =1

Country	Model	mlag+d	H_{01}	H_{02}
			?^{2}-statistic	?^{2}-statistic
Australia	2	3	7.5968[b]	2.3870
	3	2	1.1606	0.7140
Austria	2	2	5.9497[b]	0.0393
	3	5	6.9236	6.8715
Belgium	2	2	6.3536[b]	1.3320
	3	2	5.7793[b]	0.4748
Canada	2	2	0.5615	2.7076[c]
	3	3	0.1342	7.2496[b]
Denmark	2	2	1.3412	1.3270
	3	5	10.9192[b]	8.4150[c]
Finland	2	4	9.9991[b]	21.4517[a]
	3	4	8.9388[b]	18.4092[a]
	4	4	0.1770	33.0528[a]
France	2	3	1.5998	1.4826
	3	3	1.0567	0.7581
Greece	2	2	1.1844	1.0919
	3	2	1.4301	1.3893
Hungary	2	4	1.0201	5.0647
	3	4	7.1723[c]	5.6916
Iceland	2	3	19.3715[a]	2.8502
	3	3	19.0895[a]	2.5874
Ireland	2	3	5.3567[c]	0.6293
	3	2	3.2094[c]	0.1956

Italy	2	2	0.9275	3.5301[c]
	3	2	0.6169	2.7260[c]
Japan	2	4	1.5639	13.5151[a]
	3	4	2.0404	11.9566[a]
Korea Rep.	2	2	0.3941	4.7157[b]
	3	2	0.2003	5.4504[b]
Luxembourg	2	2	0.2515	2.0014
	3	3	0.2958	2.8331
Mexico	2	3	4.3251	5.8161[b]
	3	4	1.4984	1.8197
Netherlands	2	5	1.5956	5.4897
	3	3	2.4505	2.4284
New Zealand	2	4	14.5411[a]	9.4648[b]
	3	3	1.6709	7.4979[b]
Norway	2	5	2.3428	5.0255
	3	3	0.3153	0.2158
Portugal	2	5	4.5786	8.0282[c]
	3	5	4.0009	12.1542[b]
	4	5	3.5613	15.2868[a]
Spain	2	5	12.1253[b]	3.7507
	3	5	8.2080[c]	1.8065
Sweden	2	3	10.8526[a]	22.0174[a]
	3	2	8.0644[a]	9.5619[a]
Switzerland	2	5	11.0421[b]	5.7425
	3	4	10.4000[b]	5.4871
UK	2	4	11.4367[a]	19.5586[a]
	3	4	11.6854[a]	17.1836[a]

USA	2 3	3 3	1.6636 1.1104	8.6705[b] 4.5723

Bivariate systems, like the ones we have used so far, are often criticised as incomplete, omitting potentially important variables. For this reason, we consider a trivariate system as well, namely LNGDP-LNEXP-LNOPEN, where openness is expected to appraise the sensitivity of GDP and exports to each other.[19] We use $d = 2$ uniformly on all countries and, since the MWald test requires $mlag \exists d$, we experiment with $mlag + d$ = 4, 5, 6.[20]

It is important to realise, however, that in the trivariate system our analysis is partial and experimental, at best, for two reasons. Firstly, openness is treated as an auxiliary variable, i.e. it is not directly involved in the MWald tests. Consequently, we can study only direct, one-period-ahead causality between exports and economic growth, disregarding the possibility of indirect causality at longer time horizons. Since, unlike in a bivariate system, in a trivariate system no-causality for one period ahead does not imply no-causality for two or more periods ahead; the bivariate and trivariate causality test results are not really comparable. Secondly, having at most 39 observations for each variable and a maximal lag length of 6 years, the usable sample size is only 33, while each equation of the trivariate VAR has 19-21 unknown parameters.[21]

For these reasons, the results, which are not reported here, must be treated with great care and we can draw only tentative conclusions from

[19] The list of potentially important variables is 'endless', including, for example, employment, capital formation, terms of trade, total OECD output and imports. We have chosen trade openness, because it seems to have a permanent, statistically significant and economically substantial impact on economic growth. See e.g. Karras (2003).

[20] This choice is safe, but not parsimonious because according to the unit-root test results a second unit root in LNOPEN is an unlikely option for almost all countries.

[21] There are $3 \times 6 = 18$ slope parameters belonging to the lagged LNGDP and LNEXP terms, a constant and, depending on the order of the time trend, the slope parameters of t and t^2.

them. Nevertheless, it is worth to mention, that in the case of thirteen countries (Australia, Austria, Canada, Denmark, Hungary, Iceland, Ireland, Japan, Norway, Portugal, Spain, Switzerland and the USA) the MWald test in the trivariate system provides definitely more support to causality between LNGDP and LNEXP than in the bivariate system, and there are only two examples (Belgium and Korea) for the opposite. This is a surprising outcome since, due to the potentially important omitted variables, the bivariate system is expected to produce more spurious causality.

5. Concluding remarks

In this article we reported the unit-root, cointegration and MWald test results that were left out of to Kónya (2004) for brevity. First, we studied the uni- and bivariate time-series properties of the data and, building on these characteristics, employed Wald tests on appropriate parameter restrictions in bivariate VAR models in levels and/or first differences. The obvious weakness of this approach, namely its dependence on pre-testing, was clearly demonstrated by our results. Unfortunately, different unit-root and cointegration tests, and also different model specifications, can and often do lead to contradicting results. This shortcoming certainly does not come to light when applied researchers, without any serious reason, place all their faith in a single method or model. However, we used five unit-root and three cointegration tests, on various specifications.

The results were often ambiguous, so testing for causality in a single model was unjustified. We distinguished four cases, namely when LNGDP and LNEXP are I(0)-I(0), I(0)-I(1) or I(1)-I(0), I(1)-I(1) but not CI(1,1), and CI(1,1). For most of the countries we considered at least two of these possibilities and experimented with different deterministic trends. Following this indirect approach, we also applied the modified Wald test of Toda and Yamamoto (1995) which does not rely so heavily on pre-testing, but performs better in larger samples.

There are only eight countries (Canada, Iceland, Japan, Korea, Luxembourg, the Netherlands, Sweden and the UK) where different methods and specifications led to unanimous conclusions. For all other countries the causality test results are mixed. This is partly due to the fact that in most cases the true time-series properties of the data could not be discovered beyond doubt. Still, this fact would not cause much difficulty, if the causality test results were invariant to the methods applied. However, in many cases, different strategies delivered different outcomes. The other reason for ambiguity is the uncertainty regarding the deterministic trend degree. What type of a deterministic trend should be included and what is its impact on Granger causality? For example, models with and without a linear time trend often produce different causality test results. Although one could expect a time trend to act as a proxy for omitted economic variables, thus decreasing the chance of spurious causality, there are counterexamples as well. All things considered, it is clear that the causality results are usually not robust to method and specification, so their interpretation calls for great care.

In the light of this limitation, we can arrive at the following conclusions. We are confident to claim that there is no causality (NC) between exports and growth in Luxembourg and in the Netherlands, export causes growth (ECG) in Iceland, growth causes export (GCE) in Canada, Japan and Korea, and there is two-way causality (TWC) between export and growth in Sweden and in the UK. There is probably NC also in Denmark, France, Greece, Hungary and Norway, ECG in Australia, Austria and Ireland, and GCE in Finland, Portugal and the USA. Finally, in the case of Belgium, Italy, Mexico, New Zealand, Spain and Switzerland the results are too controversial to make a simple choice.

Bibliography

Dickey, D. and Fuller, W.A. (1979). 'Distribution of the Estimators for Autoregressive Time Series with a Unit Root', *Journal of the American Statistical Association*, vol. 74, pp. 427-31.

Dickey, D. and Fuller, W.A. (1981). 'Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root', *Econometrica*, vol. 49, pp. 1057-72.

Dickey, D. and Pantula, S. (1987). ‘Determining the Order of Differencing in Autoregressive Processes’, *Journal of Business and Economic Statistics*, vol. 15, pp. 455-61.

Dolado, J.L. and Lütkepohl, H. (1996). ‘Making Wald Tests Work for Cointegrated VAR Systems’, *Econometric Reviews*, vol. 15, pp. 369-86.

Dufour, J.M. and Renault, E. (1998). ‘Short Run and Long Run Causality in Time Series: Theory’, *Econometrica*, vol. 66, pp. 1099-125.

Enders, W. (1995). *Applied Econometric Time Series*, John Wiley & Sons.

Engle, R.F. and Granger, C.W.J. (1987). ‘Co-integration and Error Correction: Representation, Estimation, and Testing’, *Econometrica*, vol. 55, pp. 251-76.

Giles, J.A. and Mirza, S. (1999). ‘Some Pretesting Issues on Testing for Granger Noncausality’, *Econometrics Working Paper*, EWP9914, University of Victoria, p. 49.

Guisan, M.C. (2001): Causality and Cointegration between Consumption and GDP in 25 OECD Countries: Limitations of the Cointegration Approach, *Applied Econometrics and International Development*, vol. 1-1, pp. 1-14.

Haldrup, N. (1998). ‘An Econometric Analysis of I(2) Variables’, *Journal of Economic Surveys*, vol. 12, pp. 595-649.

Johansen, S. and Juselius, K. (1990): Maximum Likelihood Estimation and Inference on Cointegration - with Applications to the Demand for Money, *Oxford Bulletin of Economics and Statistics*, vol.52, pp. 169-210.

Hasza, D.P. and Fuller, W.A. (1979). ‘Estimation of Autoregressive Processes with Unit Roots’, *Annals of Statistics*, vol. 7, pp. 1106-20.

Karras, G. (2003): Trade Openness and Economic Growth. Can We

Estimate the Precise Effect?, *Applied Econometrics and International Development,* vol. 3, pp. 7-25.

Kónya, L. (2004): Export-Led Growth, Growth-Driven Export, Both or None? Granger Causality Analysis on OECD Countries, *Applied Econometrics and International Development*, vol. 4-1, pp. 73-94.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992). 'Testing the Null Hypothesis of Stationarity against the Alternative of a Unit Root', *Journal of Econometrics*, vol. 54, pp. 159-78.

Maddala, G.S. and Kim, I-M. (1998). *Unit Roots, Cointegration, and Structural Change*, Cambridge University Press.

Park, H.J. and Fuller, W.A. (1995). 'Alternative Estimators and Unit Root Tests for the Autoregressive Process', *Journal of Time Series Analysis*, vol. 16, pp. 415-29.

Rambaldi, A. and Doran, H. (1996). 'Testing for Granger Non-Causality in Cointegrated Systems Made Easy', *Working Papers in Econometrics and Applied Statistics*, Department of Econometrics, University of New England, No. 88, p. 23.

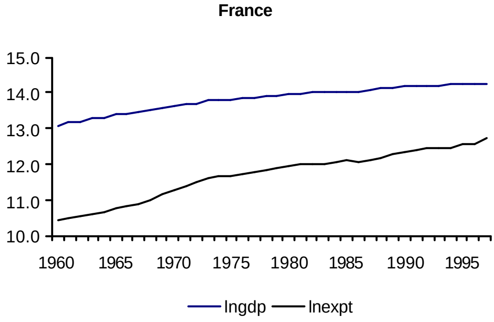
Shin, Y. (1994). 'A Residual-Based Test of the Null of Cointegration Against the Alternative of No Cointegration', *Econometric Theory*, vol. 10, pp. 91-115.

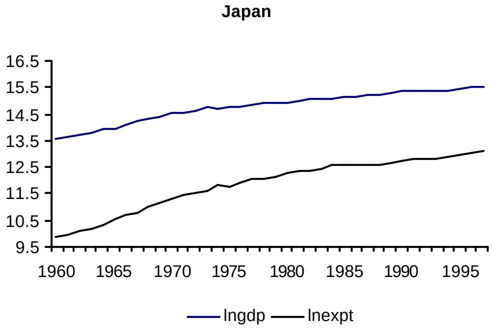
Toda, H.Y. and Yamamoto, T. (1995). 'Statistical Inference in Vector Autoregressions with Possibly Integrated Processes', *Journal of Econometrics*, vol. 66, pp. 225-50.

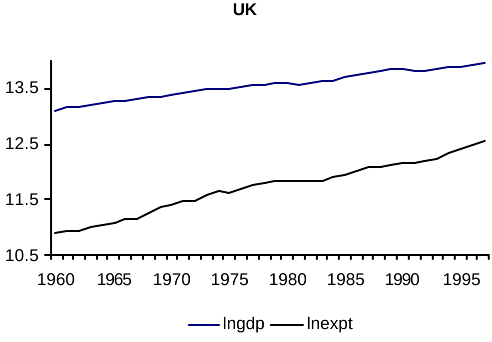
Zapata, H.O. and Rambaldi, A.N. (1997). 'Monte Carlo Evidence on Cointegration and Causation', *Oxford Bulletin of Economics and Statistics*, vol. 59, pp. 285-98.

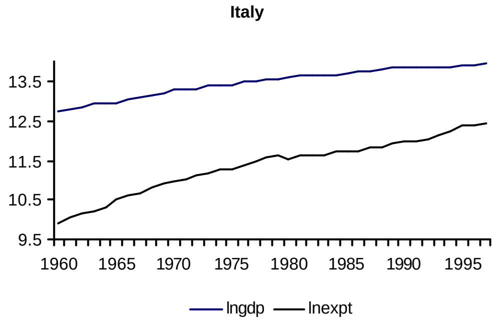
Appendix.

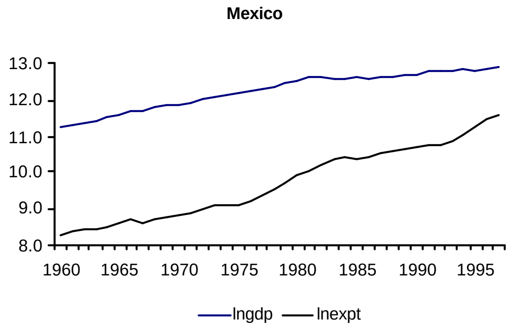
Figures Logarithms of real GDP (LNGDP) and real exports (LNEXP

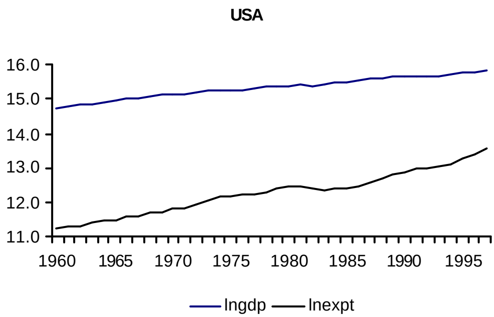







Italy
13.5
12.5
11.5
10.5
9.5
1960
1965
1970
1975
1980
1985
1990
1995
lngdp
lnexpt


Mexico
13.0
12.0
11.0
10.0
9.0
8.0
1960
1965
1970
1975
1980
1985
1990
1995
lngdp
lnexpt


USA
16.0
15.0
14.0
13.0
12.0
11.0
1960
1965
1970
1975
1980
1985
1990
1995
lngdp
lnexpt